

Poster II-11

Data Management and Data Sharing in a Modern Genetics Laboratory

Joshi, R., Wang, R., Sawant, P., Vasanth, A., Geraghty, D.E.

Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Many of the smaller laboratories that produce genetics information are faced with two major problems. The first is organizing data for examination and analysis, including cost and quality management. Sequencing technology now allows for the rapid acquisition of large amounts of genetics data by a small team of researchers, creating a need to track data quality in real time and to manage costs appropriately in order to remain competitive for research dollars. However, small labs cannot afford to develop the needed informatics infrastructure on their own. A second limitation is presented by the need for labs to share or exchange data with other labs in a secure manner with an appropriate level of control over their data, at least until such time that they are ready to publish. To effectively share genetics data, and more specifically sequence and sequencer-derived data, raw data often needs to be examined for comparison and verification. The difficulties associated with accessing such data among remote sites suggest unique informatics solutions are necessary.

Over the past 2 years, we have been involved in establishing informatics infrastructure for our efforts resequencing human immune response genes. From this experience, we have built a genetics management system (GeMS) that is applicable in a general way to the management of data acquired from automated DNA sequencers. This informatics resource includes support for three types of data acquisition, 1) PCR resequencing commonly used for SNP discovery and genotyping, 2) shotgun sequencing projects, applicable to primary data acquisition as well as resequencing of target regions from cohorts in disease studies, and 3) genotyping methods that use sequence-based data acquisition (e.g. micro satellite, SSP-SNP, HLA typing). We are currently testing this software in partner laboratories with two goals in mind. First, by formally gathering user requirements and enhancement requests with these laboratories over a period of 6 months, we are identifying new content and feature elements that increase the system's utility in a lab setting. Second, the common data format available from installation of GeMS has facilitated the ability for users to share data in a secure manner under the control of the data owner. In this regard, we are developing software that utilizes this format and allows individuals to share and exchange data between remote sites. This will support the secure sharing of DNA sequencing data between labs, in a manner that is compatible with other global initiatives in this area. We expect that this software can serve as a model for a wide variety of biological data acquisition and sharing schemes.

This work was supported by U24AI 49213-01 project 7 (D.E.G.).